

A new kernel estimator of hazard ratio and its asymptotic mean squared error

Taku MORIYAMA
 Graduate School of Mathematics
 Yoshihiko MAESONO
 Faculty of Mathematics

Kyushu University, Motooka 744, Fukuoka 819-0395, Japan

Abstract

The hazard function is a ratio of a density and survival function, and it is a basic tool of the survival analysis. In this paper we propose a kernel estimator of the hazard ratio function, which are based on a modification of Ćwik and Mielniczuk [2]’s method. We study nonparametric estimators of the hazard function and compare those estimators by means of asymptotic mean squared error ($AMSE$). We obtain asymptotic bias and variance of the new estimator, and compare them with a naive estimator. The asymptotic variance of the new estimator is always smaller than the naive estimator’s, so we also discuss an improvement of $AMSE$ using Terrell and Scott [10]’s bias reduction method. The new modified estimator ensures the non-negativity, and we demonstrate the numerical improvement.

Keywords: *Kernel estimator; hazard ratio; nonparametric estimator; mean squared error*

1 Introduction

Rosenblatt [8] has proposed a kernel estimator of a probability density function $f(\cdot)$. After that, many researchers have discussed various kernel estimators for distribution, regression, hazard functions etc. Since most of the kernel type estimators are biased, many papers have studied bias reductions. Those are based on a higher order kernel, transformations of estimators etc. Terrell and Scott [10] have proposed efficient transformation. We will apply their method to the kernel estimator in Section 4. On the other hand, Ćwik and Mielniczuk [2] have discussed a new kernel estimator of a density ratio which they called ‘direct’ estimator. Asymptotic mean squared error ($AMSE$) of the direct estimator is

different from a naive estimator, and the variance is always smaller. Here let us introduce their direct estimator.

Let X_1, X_2, \dots, X_n be independently and identically distributed (*i.i.d.*) random variables with a distribution function $F(\cdot)$, and Y_1, Y_2, \dots, Y_n be *i.i.d.* random variables with a distribution function $G(\cdot)$. $f(\cdot)$ and $g(\cdot)$ are the density functions of $F(\cdot)$ and $G(\cdot)$, and we assume $g(x_0) \neq 0$ ($x_0 \in \mathbf{R}$). A naive estimator of $f(x_0)/g(x_0)$ is defined as replacing $f(x_0)$ and $g(x_0)$ by the kernel density estimators $\hat{f}(x_0)$ and $\hat{g}(x_0)$ where

$$\hat{f}(x_0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right) \quad \text{and} \quad \hat{g}(x_0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_0 - Y_i}{h}\right).$$

$K(\cdot)$ is a kernel function and h is a bandwidth which satisfies $h \rightarrow 0$ and $nh \rightarrow \infty$ ($n \rightarrow \infty$). We call $\hat{f}(x_0)/\hat{g}(x_0)$ ‘indirect’ estimator. The direct type estimator has been proposed by Ćwik and Mielniczuk [2], which is given by

$$\frac{\hat{f}}{\hat{g}}(x_0) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{G_n(x_0) - G_n(y)}{h}\right) dF_n(y),$$

where $F_n(\cdot)$ and $G_n(\cdot)$ are the empirical distribution functions of X_1, \dots, X_n and Y_1, \dots, Y_n , respectively. Chen, Hsu, and Liaw [1] have obtained an explicit form of its *AMSE*.

Extending their method, we propose a new ‘direct’ estimator of the hazard ratio function in Section 2 and investigate its *AMSE*. We will compare the naive and direct kernel estimators in Section 3. It is demonstrated that our direct estimator performs asymptotically better, especially when the underlying distribution is exponential or gamma. Though the bias term of the direct estimator is large in some cases, the asymptotic variance is always smaller. As mentioned before, there are many bias reduction methods and then we discuss the bias reduction in Section 4. Applying the Terrell and Scott [10]’s method, we will show that the modified direct estimator gives us good performance both theoretically and numerically. Proofs of them are given in Appendices.

2 Hazard ratio estimators and asymptotic properties

2.1 Hazard ratio estimators

Let us assume that the density function $f(\cdot)$ of X_i satisfies $f(x_0) \neq 0$ ($x_0 \in \mathbf{R}$). The hazard ratio function is defined as

$$H(x) = \frac{f(x)}{1 - F(x)}.$$

The meaning of $H(x)dx$ is conditional probability of ‘death’ in $[x, x+dx]$ given a survival to x . This is a fundamental measure in survival analysis, and the range

of applications is wide. For example, actuaries call it “force of mortality” and use it to estimate insurance money. In extreme value theory, it is called “intensity function” and used to determine the form of an extreme value distribution (see Gumbel [3])

The naive nonparametric estimator of $H(x_0)$ is given by Watson [12]

$$\tilde{H}(x_0) = \frac{\hat{f}(x_0)}{1 - \hat{F}(x_0)},$$

where

$$\hat{f}(x_0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right) \quad \text{and} \quad \hat{F}(x_0) = \frac{1}{n} \sum_{i=1}^n W\left(\frac{x_0 - X_i}{h}\right).$$

Here $K(\cdot)$ is the kernel function and $W(\cdot)$ is an integral of $K(\cdot)$

$$W(u) = \int_{-\infty}^u K(t)dt.$$

Using properties of the kernel density estimator, Murthy [5] proves the consistency and asymptotic normality of $\tilde{H}(x_0)$. In random censorship model, Tanner, Wong et al. [9] proves them, using the Hájek’s projection method. Patil [6] gives its mean integrated squared error (*MISE*) and discusses the optimal bandwidth in both uncensored and censored settings. For dependent data, *MSE* of the indirect estimator is obtained by Quintela-del Río [7]. Using Vieu [11]’s results, he has obtained modified *MISE* which avoids that the denominator equal to 0. In this paper, we assume that the support of the kernel $K(\cdot)$ is $[-d, d]$ ($d > 0$) and the data do not have a censoring.

Extending the idea of Ćwik and Mielniczuk [2], we propose a new ‘direct’ estimator of the hazard ratio function as follows:

$$\hat{H}(x_0) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{y - \eta_n(y) - (x_0 - \eta_n(x_0))}{h}\right) dF_n(y),$$

where

$$\eta_n(y) = \int_{-\infty}^y F_n(u)du = n^{-1} \sum I(X_i \leq y)(y - X_i).$$

It is easy to see that $\hat{H}(x)$ is a smooth function. We will discuss its asymptotic properties.

2.2 Asymptotic properties

For brevity, we use a notation

$$A_{i,j} = \int_{-\infty}^{\infty} K^i(u)u^j du.$$

The proofs of the following theorems are given in Appendix. For the direct hazard estimator, we have the following *AMSE*.

Theorem 1 Let us assume that (i) $f(\cdot)$ is 4 times continuously differentiable at x_0 , (ii) K is symmetric and the support is $[-d, d]$ for some positive number $d > 0$ and (iii) $A_{1,4}$ and $A_{2,0}$ are bounded. Then, the MSE of $\hat{H}(x_0)$ is given by

$$\begin{aligned} & E \left\{ \hat{H}(x_0) - \left[\frac{f}{1-F} \right] (x_0) \right\}^2 \\ &= \frac{h^4}{4} A_{1,2}^2 \left[\frac{\{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3\}^2}{(1-F)^{10}} \right] (x_0) + \frac{A_{2,0}}{nh} \left[\frac{f}{1-F} \right] (x_0) \\ &+ O \left(h^6 + \frac{1}{nh^{1/2}} \right). \end{aligned}$$

On the other hand, under some regularity conditions, MSE of $\tilde{H}(x_0)$ is given by Patil [6] as follows:

$$\begin{aligned} & E \left\{ \tilde{H}(x_0) - \left[\frac{f}{1-F} \right] (x_0) \right\}^2 \\ &= \frac{h^4}{4} A_{1,2}^2 \left[\frac{\{(1-F)f'' + ff'\}^2}{(1-F)^4} \right] (x_0) + \frac{A_{2,0}}{nh} \left[\frac{f}{(1-F)^2} \right] (x_0) \quad (2) \\ &+ O \left(h^6 + \frac{1}{nh^{1/2}} \right). \end{aligned}$$

The asymptotic variances are second terms of (1) and (2), and then the direct estimator has small variance because of $0 < 1-F(x_0) < 1$. Taking the derivative of the AMSE with respect to h , we have an optimal bandwidth $h = h^*$ of $\hat{H}(x_0)$, where

$$h^* = n^{-1/5} \left(\frac{A_{2,0}}{A_{1,2}^2} \left[\frac{(1-F)^9 f}{\{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3\}^2} \right] (x_0) \right)^{1/5}.$$

Though h^* depends on unknown functions, it is available by replacing them by their estimators (plug-in method). Similarly, the following optimal bandwidth of the indirect $\tilde{H}(x_0)$ is obtained

$$h^{**} = n^{-1/5} \left(\frac{A_{2,0}}{A_{1,2}^2} \left[\frac{(1-F)^2 f}{\{(1-F)f'' + ff'\}^2} \right] (x_0) \right)^{1/5}.$$

Further we can show the asymptotic normality of the direct \hat{H} .

Theorem 2 Let us assume the same conditions of Theorem 1 and $h = cn^{-d}$ ($0 < c, \frac{1}{5} \leq d < \frac{1}{2}$). Then, the following asymptotic normality of $\hat{H}(x_0)$ holds.

$$\sqrt{nh} \left\{ \hat{H}(x_0) - \left[\frac{f}{1-F} \right] (x_0) - h^2 B_1 \right\} \xrightarrow{d} N(0, V_1),$$

where

$$B_1 = \frac{A_{1,2}}{2} \left[\frac{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3}{(1-F)^5} \right] (x_0) \quad \text{and} \quad V_1 = A_{2,0} \left[\frac{f}{1-F} \right] (x_0).$$

Using the Slutsky's theorem, the asymptotic normality of the indirect estimator is easily obtained.

The direct estimator is superior in the sense of the asymptotic variance, so we will consider the bias reduction in Section 4. We have the following higher-order asymptotic bias.

Theorem 3 *Let us assume that (i') f is 6 times continuously differentiable at x_0 , (ii') K is symmetric and the support is $[-d, d]$ for some positive number $d > 0$ and (iii') $A_{1,6}$ is bounded. Then, the higher-order asymptotic bias of $\hat{H}(x_0)$ is*

$$E \left\{ \hat{H}(x_0) - \left[\frac{f}{1-F} \right] (x_0) \right\} = h^2 a_2(x_0) + h^4 a_4(x_0) + O(h^6 + n^{-1}),$$

where

$$\begin{aligned} a_2(x_0) &= \frac{A_{1,2}}{2} \left[\frac{-m^2 m''' + 4m m' m'' - 3(m')^3}{m^5} \right] (x_0), \\ a_4(x_0) &= \frac{A_{1,4}}{24} \left[\frac{-60m^2 (m')^2 m''' + 15m^3 m'' m''' + 11m^3 m' m^{(4)} - m^4 m^{(5)}}{m^9} \right. \\ &\quad \left. + \frac{210m (m')^3 m'' - 73m^2 m' (m'')^2 - 105(m')^5}{m^9} \right] (x_0) \end{aligned}$$

and $m(x) = 1 - F(x)$.

3 Comparison of kernel hazard estimators

In some special cases, we will investigate the $AMSE$ values of the direct $\hat{H}(x_0)$ and indirect $\tilde{H}(x_0)$. We show that the proposed estimator $\hat{H}(x_0)$ performs asymptotically better when $F(\cdot)$ is the exponential or gamma distribution.

Here we consider that $F(\cdot)$ is the exponential, uniform, gamma, weibull or beta distribution. The cumulative distribution function of the exponential distribution $Exp(1/\lambda)$ is $F(x) = 1 - \exp(-\lambda x)$ and the hazard ratio is constant, that is $H(x) = \lambda$. It is one of the most common and an important model of the survival analysis. When $F(\cdot)$ is the exponential, the asymptotic biases of $\tilde{H}(x_0)$ and $\hat{H}(x_0)$ vanish and the $AMSE$ s are

$$AMSE \left[\hat{H}(x_0) \right] = \frac{\lambda}{nh} A_{2,0} < AMSE \left[\tilde{H}(x_0) \right] = \frac{\lambda}{nh} \exp(\lambda x_0) A_{2,0}.$$

Thus the proposed estimator is always asymptotically better regardless of the parameter λ and the point x_0 .

Next, let us assume that $F(\cdot)$ is the uniform distribution ($F(x) = x/b$ ($0 < x < b$)), then the hazard ratio is $H(x) = (b - x)^{-1}$. In this model, the hazard ratio increases drastically in the tail area. These *AMSEs* are the followings

$$\begin{aligned} AMSE \left[\hat{H}(x_0) \right] &= \frac{h^4}{4} A_{1,2}^2 \frac{9b^4}{(b - x_0)^{10}} + \frac{1}{nh} \frac{1}{b - x_0} A_{2,0} \\ AMSE \left[\tilde{H}(x_0) \right] &= \frac{1}{nh} \frac{b}{(b - x_0)^2} A_{2,0}. \end{aligned}$$

We find that the asymptotic bias of $\tilde{H}(x_0)$ vanishes and the variance of $\hat{H}(x_0)$ is smaller. Their asymptotic performance depends on x_0 and b , but *AMSE* of the proposed $\tilde{H}(x_0)$ is smaller when the life span b is large.

Last, we consider that F is the gamma $\Gamma(p, 100)$, weibull $W(q, 100)$ or beta distribution $100 \times B(r, s)$ where p, q, r and s are their shape parameters. Their scale $\sigma = 100$ is moderate as life span. $\Gamma(p, \sigma)$ is the distribution of a sum of p *i.i.d.* random variables of $Exp(\sigma)$, and so it is most important case. Its asymptotic squared bias, variance and *AMSE* values for some fixed points x_0 are found in Table 1, where we omit power terms of h . $\hat{\gamma}_\varepsilon^{[p]}$ and $\tilde{\gamma}_\varepsilon^{[p]}$ represent those values of $\hat{H}(x_0)$ and $\tilde{H}(x_0)$ when x_0 is ε -th quantile of $\Gamma(p, 100)$. The kernel is the Epanechnikov with $A_{1,2} = 1/5, A_{2,0} = 3/10$ and $h = n^{-1/5}$. The coefficients $n^{-4/5}$ are omitted here.

The weibull distribution $W(q, \sigma)$ is also important in survival analysis because the hazard ratio is proportional to polynomial degree $(q - 1)$, that is $H(x) = q\sigma^q x^{q-1}$. $W(1, \sigma)$ is the exponential distribution. The beta distribution is often used to describe distribution whose support is finite, and it has plentiful shapes. Table 2, 3 and 4 give each of the least *AMSE* values using h^* or h^{**} (in Section 2.2) where $w_\varepsilon^{[q]}$ and $\beta_\varepsilon^{[r,s]}$ stand *AMSE* values of $\hat{H}(x_0)$ and $\tilde{H}(x_0)$ when x_0 is ε -th quantile of $W(q, 100)$ or $B(r, s) \times 100$. We demonstrate the proposed estimator \tilde{H} performs asymptotically better in most cases of the gamma $\Gamma(p, 100)$. In the weibull cases, the asymptotic performance of our estimator is good and comparable in the beta cases.

4 Bias reduction and simulation study

As discussed in Section 3, the direct estimator has good performance, especially small variance. If we could reduce the bias, we can get a better estimator. There are many bias reduction methods, but the variance reduction is quite difficult. In this section, we discuss reduction of the bias of $\tilde{H}(x_0)$. The bias term is complicated, but if we use 4-*th* order kernel, we have $A_{1,2} = 0$. Thus we can reduce the convergence order of the bias from $O(h^2)$ to $O(h^4)$. A simple way to construct the 4-*th* order kernel is proposed by Jones [4]. However the estimator takes negative value in some cases, though the hazard ratio is always non-negative. On the other hand, Terrell and Scott [10]'s method reduces asymptotic bias without loss of non-negativity. We propose the following modified direct

Table 1: Asymptotic $Bias^2$, Var and $AMSE$ when F is gamma

	$Bias^2$	Var	$AMSE$		$Bias^2$	Var	$AMSE$
$\hat{\gamma}_{0.05}^{[1/2]}$	7.22×10^{-2}	4.01×10^{-2}	0.112	$\hat{\gamma}_{0.1}^{[1/2]}$	8.24×10^{-5}	2.11×10^{-2}	2.11×10^{-2}
$\tilde{\gamma}_{0.05}^{[1/2]}$	6.53×10^{-2}	4.22×10^{-2}	0.107	$\tilde{\gamma}_{0.1}^{[1/2]}$	6.72×10^{-5}	2.33×10^{-2}	2.34×10^{-2}
$\hat{\gamma}_{0.25}^{[1/2]}$	1.33×10^{-8}	9.52×10^{-3}	9.52×10^{-3}	$\hat{\gamma}_{0.5}^{[1/2]}$	2.36×10^{-11}	5.65×10^{-3}	5.65×10^{-3}
$\tilde{\gamma}_{0.25}^{[1/2]}$	7.74×10^{-9}	1.27×10^{-2}	1.27×10^{-2}	$\tilde{\gamma}_{0.5}^{[1/2]}$	6.83×10^{-12}	1.13×10^{-2}	1.13×10^{-2}
$\hat{\gamma}_{0.75}^{[1/2]}$	5.35×10^{-13}	4.29×10^{-3}	4.29×10^{-3}	$\hat{\gamma}_{0.9}^{[1/2]}$	1.69×10^{-15}	3.76×10^{-3}	3.76×10^{-3}
$\tilde{\gamma}_{0.75}^{[1/2]}$	6.00×10^{-14}	1.72×10^{-2}	1.72×10^{-2}	$\tilde{\gamma}_{0.9}^{[1/2]}$	2.93×10^{-15}	3.76×10^{-2}	3.76×10^{-2}
$\hat{\gamma}_{0.95}^{[1/2]}$	1.01×10^{-12}	3.58×10^{-3}	3.58×10^{-3}	$\hat{\gamma}_{0.975}^{[1/2]}$	1.47×10^{-11}	3.46×10^{-3}	3.46×10^{-3}
$\tilde{\gamma}_{0.95}^{[1/2]}$	6.93×10^{-16}	7.16×10^{-2}	7.16×10^{-2}	$\tilde{\gamma}_{0.975}^{[1/2]}$	2.33×10^{-16}	0.139	0.139
$\hat{\gamma}_{0.05}^{[10]}$	2.49×10^{-18}	1.56×10^{-4}	1.56×10^{-4}	$\hat{\gamma}_{0.1}^{[10]}$	2.16×10^{-18}	2.55×10^{-4}	2.55×10^{-4}
$\tilde{\gamma}_{0.05}^{[10]}$	7.16×10^{-19}	1.64×10^{-4}	1.64×10^{-4}	$\tilde{\gamma}_{0.1}^{[10]}$	1.73×10^{-21}	2.83×10^{-4}	2.83×10^{-4}
$\hat{\gamma}_{0.25}^{[10]}$	2.61×10^{-18}	4.77×10^{-4}	4.77×10^{-4}	$\hat{\gamma}_{0.5}^{[10]}$	1.37×10^{-17}	7.72×10^{-4}	7.72×10^{-4}
$\tilde{\gamma}_{0.25}^{[10]}$	2.40×10^{-18}	6.36×10^{-4}	6.36×10^{-4}	$\tilde{\gamma}_{0.5}^{[10]}$	7.91×10^{-18}	1.54×10^{-3}	1.54×10^{-3}
$\hat{\gamma}_{0.75}^{[10]}$	2.84×10^{-16}	1.07×10^{-3}	1.07×10^{-3}	$\hat{\gamma}_{0.9}^{[10]}$	1.19×10^{-14}	1.32×10^{-3}	1.32×10^{-3}
$\tilde{\gamma}_{0.75}^{[10]}$	1.05×10^{-17}	4.28×10^{-3}	4.28×10^{-3}	$\tilde{\gamma}_{0.9}^{[10]}$	9.83×10^{-18}	1.32×10^{-2}	1.32×10^{-2}
$\hat{\gamma}_{0.95}^{[10]}$	1.84×10^{-13}	1.45×10^{-3}	1.45×10^{-3}	$\hat{\gamma}_{0.975}^{[10]}$	2.75×10^{-12}	1.56×10^{-3}	1.56×10^{-3}
$\tilde{\gamma}_{0.95}^{[10]}$	8.70×10^{-18}	2.90×10^{-2}	2.90×10^{-2}	$\tilde{\gamma}_{0.975}^{[10]}$	7.58×10^{-18}	6.24×10^{-2}	6.24×10^{-2}

Table 2: $AMSE$ values when F is gamma and $h = h^*$ or h^{**}

	$AMSE$		$AMSE$		$AMSE$		$AMSE$
$\hat{\gamma}_{0.05}^{[1/2]}$	7.44×10^{-2}	$\hat{\gamma}_{0.1}^{[1/2]}$	1.14×10^{-2}	$\hat{\gamma}_{0.25}^{[1/2]}$	1.06×10^{-3}	$\hat{\gamma}_{0.5}^{[1/2]}$	1.97×10^{-4}
$\tilde{\gamma}_{0.05}^{[1/2]}$	7.60×10^{-2}	$\tilde{\gamma}_{0.1}^{[1/2]}$	1.19×10^{-2}	$\tilde{\gamma}_{0.25}^{[1/2]}$	1.20×10^{-3}	$\tilde{\gamma}_{0.5}^{[1/2]}$	2.67×10^{-4}
$\hat{\gamma}_{0.75}^{[1/2]}$	7.40×10^{-5}	$\hat{\gamma}_{0.9}^{[1/2]}$	2.11×10^{-5}	$\hat{\gamma}_{0.95}^{[1/2]}$	7.26×10^{-5}	$\hat{\gamma}_{0.975}^{[1/2]}$	1.21×10^{-4}
$\tilde{\gamma}_{0.75}^{[1/2]}$	1.45×10^{-4}	$\tilde{\gamma}_{0.9}^{[1/2]}$	1.48×10^{-4}	$\tilde{\gamma}_{0.95}^{[1/2]}$	1.86×10^{-4}	$\tilde{\gamma}_{0.975}^{[1/2]}$	2.54×10^{-4}
$\hat{\gamma}_{0.05}^{[10]}$	4.48×10^{-7}	$\hat{\gamma}_{0.1}^{[10]}$	6.45×10^{-7}	$\hat{\gamma}_{0.25}^{[10]}$	1.11×10^{-6}	$\hat{\gamma}_{0.5}^{[10]}$	2.26×10^{-6}
$\tilde{\gamma}_{0.05}^{[10]}$	3.64×10^{-7}	$\tilde{\gamma}_{0.1}^{[10]}$	1.68×10^{-7}	$\tilde{\gamma}_{0.25}^{[10]}$	1.37×10^{-6}	$\tilde{\gamma}_{0.5}^{[10]}$	3.53×10^{-6}
$\hat{\gamma}_{0.75}^{[10]}$	5.39×10^{-6}	$\hat{\gamma}_{0.9}^{[10]}$	1.34×10^{-5}	$\hat{\gamma}_{0.95}^{[10]}$	2.51×10^{-5}	$\hat{\gamma}_{0.975}^{[10]}$	4.57×10^{-5}
$\tilde{\gamma}_{0.75}^{[10]}$	8.46×10^{-6}	$\tilde{\gamma}_{0.9}^{[10]}$	2.05×10^{-5}	$\tilde{\gamma}_{0.95}^{[10]}$	3.76×10^{-5}	$\tilde{\gamma}_{0.975}^{[10]}$	6.75×10^{-5}

Table 3: $AMSE$ values when F is weibull and $h = h^*$ or h^{**}

	$AMSE$		$AMSE$		$AMSE$		$AMSE$
$\hat{w}_{0.05}^{[1/2]}$	4.11×10^{-2}	$\hat{w}_{0.1}^{[1/2]}$	5.68×10^{-3}	$\hat{w}_{0.25}^{[1/2]}$	3.85×10^{-4}	$\hat{w}_{0.5}^{[1/2]}$	4.25×10^{-5}
$\tilde{w}_{0.05}^{[1/2]}$	4.20×10^{-2}	$\tilde{w}_{0.1}^{[1/2]}$	5.93×10^{-3}	$\tilde{w}_{0.25}^{[1/2]}$	4.30×10^{-4}	$\tilde{w}_{0.5}^{[1/2]}$	5.50×10^{-5}
$\hat{w}_{0.75}^{[1/2]}$	9.22×10^{-6}	$\hat{w}_{0.9}^{[1/2]}$	3.31×10^{-6}	$\hat{w}_{0.95}^{[1/2]}$	3.59×10^{-7}	$\hat{w}_{0.975}^{[1/2]}$	2.67×10^{-6}
$\tilde{w}_{0.75}^{[1/2]}$	1.53×10^{-5}	$\tilde{w}_{0.9}^{[1/2]}$	8.61×10^{-6}	$\tilde{w}_{0.95}^{[1/2]}$	7.68×10^{-6}	$\tilde{w}_{0.975}^{[1/2]}$	7.90×10^{-6}
$\hat{w}_{0.05}^{[10]}$	1.20×10^{-4}	$\hat{w}_{0.1}^{[10]}$	2.65×10^{-4}	$\hat{w}_{0.25}^{[10]}$	9.00×10^{-4}	$\hat{w}_{0.5}^{[10]}$	3.40×10^{-3}
$\tilde{w}_{0.05}^{[10]}$	1.11×10^{-4}	$\tilde{w}_{0.1}^{[10]}$	2.23×10^{-4}	$\tilde{w}_{0.25}^{[10]}$	4.79×10^{-4}	$\tilde{w}_{0.5}^{[10]}$	2.34×10^{-3}
$\hat{w}_{0.75}^{[10]}$	1.38×10^{-2}	$\hat{w}_{0.9}^{[10]}$	5.49×10^{-2}	$\hat{w}_{0.95}^{[10]}$	0.135	$\hat{w}_{0.975}^{[10]}$	0.309
$\tilde{w}_{0.75}^{[10]}$	1.33×10^{-2}	$\tilde{w}_{0.9}^{[10]}$	5.96×10^{-2}	$\tilde{w}_{0.95}^{[10]}$	0.153	$\tilde{w}_{0.975}^{[10]}$	0.360

Table 4: *AMSE* values when F is beta and $h = h^*$ or h^{**}

	<i>AMSE</i>		<i>AMSE</i>		<i>AMSE</i>		<i>AMSE</i>
$\hat{\beta}_{0.05}^{[1/2,1/2]}$	7.61×10^{-3}	$\hat{\beta}_{0.1}^{[1/2,1/2]}$	1.20×10^{-3}	$\hat{\beta}_{0.25}^{[1/2,1/2]}$	1.42×10^{-4}	$\hat{\beta}_{0.5}^{[1/2,1/2]}$	1.46×10^{-4}
$\tilde{\beta}_{0.05}^{[1/2,1/2]}$	7.77×10^{-3}	$\tilde{\beta}_{0.1}^{[1/2,1/2]}$	1.25×10^{-3}	$\tilde{\beta}_{0.25}^{[1/2,1/2]}$	1.54×10^{-4}	$\tilde{\beta}_{0.5}^{[1/2,1/2]}$	1.11×10^{-4}
$\hat{\beta}_{0.75}^{[1/2,1/2]}$	2.35×10^{-3}	$\hat{\beta}_{0.9}^{[1/2,1/2]}$	0.167	$\hat{\beta}_{0.95}^{[1/2,1/2]}$	4.57	$\hat{\beta}_{0.975}^{[1/2,1/2]}$	127
$\tilde{\beta}_{0.75}^{[1/2,1/2]}$	1.61×10^{-3}	$\tilde{\beta}_{0.9}^{[1/2,1/2]}$	0.116	$\tilde{\beta}_{0.95}^{[1/2,1/2]}$	3.17	$\tilde{\beta}_{0.975}^{[1/2,1/2]}$	87.9
$\hat{\beta}_{0.05}^{[2,5]}$	1.55×10^{-4}	$\hat{\beta}_{0.1}^{[2,5]}$	2.70×10^{-4}	$\hat{\beta}_{0.25}^{[2,5]}$	5.35×10^{-4}	$\hat{\beta}_{0.5}^{[2,5]}$	1.15×10^{-3}
$\tilde{\beta}_{0.05}^{[2,5]}$	1.83×10^{-4}	$\tilde{\beta}_{0.1}^{[2,5]}$	2.18×10^{-4}	$\tilde{\beta}_{0.25}^{[2,5]}$	2.91×10^{-4}	$\tilde{\beta}_{0.5}^{[2,5]}$	5.67×10^{-4}
$\hat{\beta}_{0.75}^{[2,5]}$	3.09×10^{-3}	$\hat{\beta}_{0.9}^{[2,5]}$	1.01×10^{-2}	$\hat{\beta}_{0.95}^{[2,5]}$	2.41×10^{-2}	$\hat{\beta}_{0.975}^{[2,5]}$	5.70×10^{-2}
$\tilde{\beta}_{0.75}^{[2,5]}$	1.84×10^{-3}	$\tilde{\beta}_{0.9}^{[2,5]}$	7.22×10^{-3}	$\tilde{\beta}_{0.95}^{[2,5]}$	1.88×10^{-2}	$\tilde{\beta}_{0.975}^{[2,5]}$	4.72×10^{-2}
$\hat{\beta}_{0.05}^{[5,2]}$	7.18×10^{-5}	$\hat{\beta}_{0.1}^{[5,2]}$	1.38×10^{-4}	$\hat{\beta}_{0.25}^{[5,2]}$	4.20×10^{-4}	$\hat{\beta}_{0.5}^{[5,2]}$	1.72×10^{-3}
$\tilde{\beta}_{0.05}^{[5,2]}$	6.57×10^{-5}	$\tilde{\beta}_{0.1}^{[5,2]}$	1.16×10^{-4}	$\tilde{\beta}_{0.25}^{[5,2]}$	2.84×10^{-4}	$\tilde{\beta}_{0.5}^{[5,2]}$	6.73×10^{-4}
$\hat{\beta}_{0.75}^{[5,2]}$	9.66×10^{-3}	$\hat{\beta}_{0.9}^{[5,2]}$	6.68×10^{-2}	$\hat{\beta}_{0.95}^{[5,2]}$	0.261	$\hat{\beta}_{0.975}^{[5,2]}$	0.981
$\tilde{\beta}_{0.75}^{[5,2]}$	4.23×10^{-3}	$\tilde{\beta}_{0.9}^{[5,2]}$	4.01×10^{-2}	$\tilde{\beta}_{0.95}^{[5,2]}$	0.171	$\tilde{\beta}_{0.975}^{[5,2]}$	0.676

hazard estimator

$$\hat{H}^I(x) = \{\hat{H}_h(x)\}^{\frac{4}{3}} \{\hat{H}_{2h}(x)\}^{-\frac{1}{3}},$$

where $\hat{H}_h(\cdot)$ and $\hat{H}_{2h}(\cdot)$ are the direct estimator with the bandwidth h and $2h$, respectively. By [10] we can get an asymptotic bias of $\hat{H}^I(x_0)$

$$\frac{2a_2^2(x_0) - 4a_4(x_0)H(x_0)}{H(x_0)}h^4,$$

where $a_2(x_0)$ and $a_4(x_0)$ are given in Theorem 2.3. Whereas, the order of the variance is

$$Var[\hat{H}^I(x_0)] = Var\left[\frac{4}{3}\hat{H}_h(x_0) - \frac{1}{3}\hat{H}_{2h}(x_0)\right] + O(n^{-1}) = O((nh)^{-1}),$$

and then the order is same as the original. The order of the optimal bandwidth is $O(h^{-1/9})$, and the optimal convergence rate is improved to $O(n^{-8/9})$. We confirm this in numerical simulations.

Table 5 shows the results of 10,000 times simulation of improvement rate (%) of

$$1 - \{MISE[\hat{H}^I]/MISE[\hat{H}]\},$$

when $F(\cdot)$ is the beta $B(r, s)$ and a sample size is $n = 100$. Their kernel functions are the Epanechnikov. The bandwidth h^\dagger of \hat{H} is the asymptotically optimal in the sense of $MISE[\hat{H}]$ which is available from Theorem 2.1 and that of \hat{H}^I is $h^{\dagger^{5/9}}$ modified to be optimal in the sense of the convergence rate. Table 5 shows that how well the bias reduction works, and $\hat{H}^I(x)$ is superior. To obtain an explicit form of the asymptotic variance $Var[\hat{H}^I(x)]$ is quite difficult and so we postpone it in future work.

Table 5: Improvement rate (%) of $MISE[\hat{H}^I] - F$ is $B(r, s) * \sigma$, $h = h^*$ or h^{**}

Improvement rate (%)	$\sigma = 4^{-1}$	$\sigma = 1$	$\sigma = 4$	$\sigma = 4^2$	$\sigma = 4^3$
$r = 1/2, s = 1/2$	0.301	1.25	1.52	1.64	1.78
$r = 1, s = 1$	0.593	1.05	1.24	1.38	1.30
$r = 3, s = 3$	6.74	8.35	8.27	8.28	8.35
$r = 2, s = 5$	12.9	13.9	15.0	14.8	15.0
$r = 5, s = 2$	3.96	5.10	5.11	5.05	5.05

References

- [1] Chen, S.M., Hsu, Y.S., and Liaw, J.T. (2009). On kernel estimators of density ratio. *Statistics* 43, 463–479.
- [2] Ćwik, J., and Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis. *Communications in Statistics-Theory and Methods*, 18, 3057–3069.
- [3] Gumbel, E. (1958). Statistics of extremes. *Columbia Univ. press, New York*.
- [4] Jones, M., and Signorini, D. (1997). A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association* 92, 1063–1073.
- [5] Murthy, V. (1965). Estimation of jumps, reliability and hazard rate. *The Annals of Mathematical Statistics*, 36, 1032–1040.
- [6] Patil, P. (1993). On the least squares cross-validation bandwidth in hazard rate estimation. *The Annals of Statistics*, 21, 1792–1810.
- [7] Quintela-del Río, A. (2007). Plug-in bandwidth selection in kernel hazard estimation from dependent data. *Computational Statistics & Data Analysis*, 51, 5800–5812.
- [8] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27, 832–837.
- [9] Tanner, M.A., Wong, W.H., et al. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *The Annals of Statistics*, 11, 989–993.
- [10] Terrell, G.R., and Scott, D.W. (1980). On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, 8, 1160–1163.
- [11] Vieu, P. (1991). Quadratic errors for nonparametric estimates under dependence. *Journal of Multivariate Analysis*, 39, 324–347.
- [12] Watson, G., and Leadbetter, M. (1964). Hazard analysis II. *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 101–116.

5 Appendices: Proofs of Theorems

Proof of Theorem 2.1

For the sake of simplicity, we use the following notations

$$\begin{aligned}\eta(z) &= \int_{-\infty}^z F(u)du, & \gamma(z) &= \int_{-\infty}^z \eta(u)du, \\ M(z) &= z - \eta(z), & m(z) &= M'(z) = 1 - F(z).\end{aligned}$$

To begin with, we consider the following stochastic expansion of the direct estimator

$$\begin{aligned}& \hat{H}(x_0) \\&= \frac{1}{h} \int K \left(\frac{M(y) - M(x_0)}{h} \right) dF_n(y) \\&+ \frac{1}{h^2} \int K' \left(\frac{M(y) - M(x_0)}{h} \right) \{[\eta(y) - \eta_n(y)] - [\eta(x_0) - \eta_n(x_0)]\} dF_n(y) \\&+ \frac{1}{h^3} \int K'' \left(\frac{M(y) - M(x_0)}{h} \right) \{[\eta(y) - \eta_n(y)] - [\eta(x_0) - \eta_n(x_0)]\}^2 dF_n(y) \\&+ \cdots \\&= J_1 + J_2 + J_3 + \cdots \quad (\text{say}).\end{aligned}$$

Since J_1 is a sum of *i.i.d.* random variables, the expectation is obtained directly

$$\begin{aligned}E[J_1] &= E \left[\frac{1}{h} \int K \left(\frac{M(y) - M(x_0)}{h} \right) dF_n(y) \right] \\&= \frac{1}{h} \int K \left(\frac{M(y) - M(x_0)}{h} \right) f(y) dy \\&= \int K(u) \left[\frac{f}{1-F} \right] (M^{-1}(M(x_0) + hu)) du \\&= \left[\frac{f}{1-F} \right] (x_0) + \frac{h^2}{2} A_{1,2} \left[\frac{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3}{(1-F)^5} \right] (x_0) + O(h^4).\end{aligned}$$

Combining the following second moment

$$\begin{aligned}& \frac{1}{nh^2} \int K^2 \left(\frac{M(y) - M(x_0)}{h} \right) f(y) dy \\&= \frac{1}{nh} \int K^2(u) \left[\frac{f}{1-F} \right] (M^{-1}(M(x_0) + hu)) du \\&= \frac{1}{nh} \frac{f}{1-F} (x_0) A_{2,0} + O(n^{-1}),\end{aligned}$$

we have the variance as follows

$$V[J_1] = \frac{1}{nh} \left[\frac{f}{1-F} \right] (x_0) A_{2,0} + O(n^{-1}).$$

Next, we consider the following representation of J_2

$$J_2 = \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{j=1}^n K' \left(\frac{M(X_i) - M(x_0)}{h} \right) Q(X_i, X_j),$$

where

$$Q(x_i, x_j) = [\eta(x_i) - J(x_i - x_j)] - [\eta(x_0) - J(x_0 - x_j)]$$

and

$$J(y - X_i) = I(X_i \leq y)(y - X_i).$$

Using the conditional expectation, we get the following equation

$$\begin{aligned} E[J_2] &= \frac{1}{nh^2} \sum_{j=1}^n E \left[K' \left(\frac{M(X_i) - M(x_0)}{h} \right) Q(X_i, X_j) \right] \\ &= \frac{1}{nh^2} E \left[K' \left(\frac{M(X_i) - M(x_0)}{h} \right) E \left[\sum_{j=1}^n Q(X_i, X_j) \mid X_i \right] \right] \\ &= \frac{1}{nh^2} E \left[K' \left(\frac{M(X_i) - M(x_0)}{h} \right) \{ \eta(X_i) - [\eta(x_0) - J(x_0 - X_i)] \} \right] \\ &= \frac{1}{nh} \int K'(u) \{ \eta(M^{-1}(M(x_0) + hu)) - \eta(x_0) + J(x_0 - M^{-1}(M(x_0) + hu)) \} \\ &\quad \times \frac{f}{1-F}(M^{-1}(M(x_0) + hu)) du \\ &= \frac{1}{nh} \int K'(u) O(hu) \frac{f}{1-F}(x_0) du = O\left(\frac{1}{n}\right). \end{aligned}$$

Next we have

$$\begin{aligned} J_2^2 &= \frac{1}{n^4 h^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n K' \left(\frac{M(X_i) - M(x_0)}{h} \right) K' \left(\frac{M(X_k) - M(x_0)}{h} \right) \\ &\quad \times Q(X_i, X_j) Q(X_k, X_l) \\ &= \frac{1}{n^4 h^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n J_2^2(i, j, k, l) \quad (\text{say}). \end{aligned}$$

Taking a conditional expectation, we obtain that if all of (i, j, k, l) are different,

$$E[J_2^2(i, j, k, l)] = E[E\{J_2^2(i, j, k, l) \mid X_i, X_k\}] = 0,$$

and

$$\begin{aligned} E[J_2^2(i, j, k, l)] &= 0 && (\text{if } i = j \text{ and all of } (i, k, l) \text{ are different}), \\ E[J_2^2(i, j, k, l)] &= 0 && (\text{if } i = k \text{ and all of } (i, j, l) \text{ are different}), \\ E[J_2^2(i, j, k, l)] &= 0 && (\text{if } i = l \text{ and all of } (i, j, k) \text{ are different}). \end{aligned}$$

Then the terms in which ‘ $j = l$ and all of (i, j, k) are different’ are the main term of $E[J_2^2]$. If $j = l$ and all of (i, j, k) are different, we have

$$\begin{aligned}
& E[J_2^2(i, j, k, l)] \\
= & \frac{n(n-1)(n-2)}{n^4 h^4} E \left[K' \left(\frac{M(X_i) - M(x_0)}{h} \right) K' \left(\frac{M(X_k) - M(x_0)}{h} \right) \right. \\
& \left. \times Q(X_i, X_j) Q(X_k, X_j) \right].
\end{aligned}$$

Using the conditional expectation of $Q(X_i, X_j)Q(X_k, X_j)$ given X_i and X_k , we find

$$\begin{aligned}
& E \left[E \left\{ Q(X_i, X_j) Q(X_k, X_j) \mid X_i, X_k \right\} \right] \\
= & E \left[\eta(X_i) \eta(x_0) + \eta(X_k) \eta(x_0) - \eta^2(x_0) + 2\gamma(x_0) - \eta(X_i) \eta(X_k) \right. \\
& - (x + X_i - 2 \min(x, X_i)) \eta(\min(x, X_i)) - 2\gamma(\min(x, X_i)) \\
& - (x + X_k - 2 \min(x, X_k)) \eta(\min(x, X_k)) - 2\gamma(\min(x, X_k)) \\
& \left. + (X_i + X_k - 2 \min(X_i, X_k)) \eta(\min(X_i, X_k)) + 2\gamma(\min(X_i, X_k)) \right].
\end{aligned}$$

Therefore, the entire expectation of the last row is

$$\begin{aligned}
& E \left[K' \left(\frac{M(X_i) - M(x_0)}{h} \right) K' \left(\frac{M(X_k) - M(x_0)}{h} \right) \right. \\
& \left. \times (X_i + X_k - 2 \min(X_i, X_k)) \eta(\min(X_i, X_k)) + 2\gamma(\min(X_i, X_k)) \right] \\
= & \int \left[\int_{-\infty}^y K' \left(\frac{M(w) - M(x_0)}{h} \right) K' \left(\frac{M(y) - M(x_0)}{h} \right) \right. \\
& \left. \times \{(-w + y) \eta(w) + 2\gamma(w)\} f(y) dw \right. \\
& + \int_y^{\infty} K' \left(\frac{M(w) - M(x_0)}{h} \right) K' \left(\frac{M(y) - M(x_0)}{h} \right) \\
& \left. \times \{(w - y) \eta(y) + 2\gamma(y)\} f(w) dw \right] f(y) dy.
\end{aligned}$$

Finally we get

$$\begin{aligned}
& E \left[K' \left(\frac{M(X_i) - M(x_0)}{h} \right) K' \left(\frac{M(X_k) - M(x_0)}{h} \right) \right. \\
& \quad \left. \times \{X_i + X_k - 2 \min(X_i, X_k)\} \eta(\min(X_i, X_k)) + 2\gamma(\min(X_i, X_k)) \right] \\
&= -h^2 \int K' \left(\frac{M(y) - M(x_0)}{h} \right) f(y) dy \\
& \quad \times \left(\left[W \left(\frac{M(y) - M(x_0)}{h} \right) \left(\{ \eta(x_0) + (-x_0 + y) F(x_0) \} \left[\frac{f}{m^2} \right] (x_0) \right. \right. \right. \\
& \quad \left. \left. \left. + \{ (-x_0 + y) \eta(x_0) + 2\gamma(x_0) \} \left[\frac{f'm - fm'}{m^3} \right] (x_0) \right) + O(h) \right] \right. \\
& \quad \left. + \left[\left(1 - W \left(\frac{M(y) - M(x_0)}{h} \right) \right) \right. \right. \\
& \quad \left. \left. \times \left(\eta(y) \frac{f}{m^2} (x_0) + \{ (x_0 - y) \eta(y) + 2\gamma(y) \} \left[\frac{f'm - fm'}{m^3} \right] (x_0) \right) + O(h) \right] \right) \\
&= h^4 \left[\frac{Ff^2}{m^4} (x_0) + \left\{ \frac{f'm - fm'}{m^3} \left(2\eta \frac{f}{m^2} + 2\gamma \frac{f'm - fm'}{m^3} \right) \right\} (x_0) \right] + O(h^5).
\end{aligned}$$

After similar calculations of the other terms, if $j = l$ and all of (i, j, k) are different, we get

$$E[J_2^2(i, j, k, l)] = O\left(\frac{1}{n}\right).$$

In addition, it is easy to see that the expectations of the other combinations of (i, j, k, l) are $o(n^{-1})$. As a result, we have

$$E[J_2^2] = O(n^{-1}) \quad \text{and} \quad V[J_2] = O(n^{-1}).$$

The moments of J_3 is also obtained in a similar manner and we find that $J_3 + \dots$ is negligible. The main bias of $\hat{H}(x_0)$ comes from J_1 . By Cauchy-Schwarz inequality, we find that the main term of the variance comes from J_1 . Now, we get the *AMSE* of the direct estimator and the proof of Theorem 2.1.

Proof of Theorem 2.2 and 2.3

It follows from the previous discussion that

$$\begin{aligned}
& \sqrt{nh} \left\{ \hat{H}(x_0) - \left[\frac{f}{1-F} \right] (x_0) \right\} \\
&= \sqrt{nh} \left\{ J_1 - \left[\frac{f}{1-F} \right] (x_0) \right\} + o_P(1) \\
&= B_1 + \sqrt{nh} \left\{ J_1 - \left[\frac{f}{1-F} \right] (x_0) - h^2 B_1 \right\} + o_P(1).
\end{aligned}$$

Since J_1 is a sum of *i.i.d.* random variables and the expectation of the second term is $o(1)$, the asymptotic normality of Theorem 2.2 holds. Furthermore, these arguments show that

$$\begin{aligned} E\left[\widehat{H}(x_0)\right] &= E[J_1] + O(n^{-1}) \\ &= \int K(u) \left[\frac{f}{1-F}\right] (M^{-1}(M(x_0) + hu)) du + O(n^{-1}), \end{aligned}$$

and we get directly Theorem 2.3 by the Taylor expansion.